

JSTOR documents released in act of solidarity with Aaron Swartz

repost 2011-07-21 20:41

This essay originally appeared at The Pirate Bay as the description for the torrent of documents. You can see this message and the torrent here: <http://thepiratebay.org/torrent/6554331>

See also Defend Internet Activist Aaron Swartz <http://sheffield.indymedia.org.uk/2011/07/482483.html>

This archive contains 18,592 scientific publications totaling 33GiB, all from Philosophical Transactions of the Royal Society and which should be available to everyone at no cost, but most have previously only been made available at high prices through paywall gatekeepers like JSTOR.

Limited access to the documents here is typically sold for \$19 USD per article, though some of the older ones are available as cheaply as \$8. Purchasing access to this collection one article at a time would cost hundreds of thousands of dollars.

Also included is the basic factual metadata allowing you to locate works by title, author, or publication date, and a checksum file to allow you to check for corruption.

I've had these files for a long time, but I've been afraid that if I published them I would be subject to unjust legal harassment by those who profit from controlling access to these works.

I now feel that I've been making the wrong decision.

On July 19th 2011, Aaron Swartz was criminally charged by the US Attorney General's office for, effectively, downloading too many academic papers from JSTOR.

Academic publishing is an odd system -- the authors are not paid for their writing, nor are the peer reviewers (they're just more unpaid academics), and in some fields even the journal editors are unpaid. Sometimes the authors must even pay the publishers.

And yet scientific publications are some of the most outrageously expensive pieces of literature you can buy. In the past, the high access fees supported the costly mechanical reproduction of niche paper journals, but online distribution has mostly made this function obsolete.

As far as I can tell, the money paid for access today serves little significant purpose except to perpetuate dead business models. The "publish or perish" pressure in academia gives the authors an impossibly weak negotiating position, and the existing system has enormous inertia.

Those with the most power to change the system--the long-tenured luminary scholars whose works give legitimacy and prestige to the journals, rather than the other way around--are the least impacted by its failures. They are supported by institutions who invisibly provide access to all of the resources they need. And as the journals depend on them, they may ask for alterations to the standard contract without risking their career on the loss of a publication offer. Many don't even realize the extent to which academic work is inaccessible to the general public, nor do they realize what sort of work is being done outside universities that would benefit by it.

Large publishers are now able to purchase the political clout needed to abuse the narrow commercial scope of copyright protection, extending it to completely inapplicable areas: slavish reproductions of historic documents and art, for example, and exploiting the labors of unpaid scientists. They're even able to make the taxpayers pay for their attacks on free society by pursuing criminal prosecution (copyright has classically been a civil matter) and by burdening public institutions with outrageous subscription fees.

Copyright is a legal fiction representing a narrow compromise: we give up some of our natural right to exchange information in exchange for creating an economic incentive to author, so that we may all enjoy more works. When publishers abuse the system to prop up their existence, when they misrepresent the extent of copyright coverage, when they use threats of frivolous litigation to suppress the dissemination of publicly owned works, they are stealing from everyone else.

Several years ago I came into possession, through rather boring and lawful means, of a large collection of JSTOR [Indymedia does blah. Content is good, and free to use for non-commercial purposes under the Open Content license. if you have questions, email someone.]

documents.

These particular documents are the historic back archives of the Philosophical Transactions of the Royal Society -- a prestigious scientific journal with a history extending back to the 1600s.

The portion of the collection included in this archive, ones published prior to 1923 and therefore obviously in the public domain, total some 18,592 papers and 33 gigabytes of data.

The documents are part of the shared heritage of all mankind, and are rightfully in the public domain, but they are not available freely. Instead the articles are available at \$19 each--for one month's viewing, by one person, on one computer. It's a steal. From you.

When I received these documents I had grand plans of uploading them to Wikipedia's sister site for reference works, Wikisource -- where they could be tightly interlinked with Wikipedia, providing interesting historical context to the encyclopedia articles. For example, Uranus was discovered in 1781 by William Herschel; why not take a look at the paper where he originally disclosed his discovery? (Or one of the several follow on publications about its satellites, or the dozens of other papers he authored?)

But I soon found the reality of the situation to be less than appealing: publishing the documents freely was likely to bring frivolous litigation from the publishers.

As in many other cases, I could expect them to claim that their slavish reproduction -- scanning the documents -- created a new copyright interest. Or that distributing the documents complete with the trivial watermarks they added constituted unlawful copying of that mark. They might even pursue strawman criminal charges claiming that whoever obtained the files must have violated some kind of anti-hacking laws.

In my discreet inquiry, I was unable to find anyone willing to cover the potentially unbounded legal costs I risked, even though the only unlawful action here is the fraudulent misuse of copyright by JSTOR and the Royal Society to withhold access from the public to that which is legally and morally everyone's property.

In the meantime, and to great fanfare as part of their 350th anniversary, the RSOL opened up "free" access to their historic archives -- but "free" only meant "with many odious terms", and access was limited to about 100 articles.

All too often journals, galleries, and museums are becoming not disseminators of knowledge -- as their lofty mission statements suggest -- but censors of knowledge, because censoring is the one thing they do better than the Internet does. Stewardship and curation are valuable functions, but their value is negative when there is only one steward and one curator, whose judgment reigns supreme as the final word on what everyone else sees and knows. If their recommendations have value they can be heeded without the coercive abuse of copyright to silence competition.

The liberal dissemination of knowledge is essential to scientific inquiry. More than in any other area, the application of restrictive copyright is inappropriate for academic works: there is no sticky question of how to pay authors or reviewers, as the publishers are already not paying them. And unlike 'mere' works of entertainment, liberal access to scientific work impacts the well-being of all mankind. Our continued survival may even depend on it.

If I can remove even one dollar of ill-gained income from a poisonous industry which acts to suppress scientific and historic understanding, then whatever personal cost I suffer will be justified -- it will be one less dollar spent in the war against knowledge. One less dollar spent lobbying for laws that make downloading too many scientific papers a crime.

I had considered releasing this collection anonymously, but others pointed out that the obviously overzealous prosecutors of Aaron Swartz would probably accuse him of it and add it to their growing list of ridiculous charges. This didn't sit well with my conscience, and I generally believe that anything worth doing is worth attaching your name to.

I'm interested in hearing about any enjoyable discoveries or even useful applications which come of this archive.

- ---- Greg Maxwell - July 20th 2011 gmaxwell@gmail.com Bitcoin: 14csFEJHk3SYbkBmajyJ3ktpsd2TmwDEBb

<http://www.indymedia.org.uk/en/2011/07/482537.shtml>